ISSN 1745-8587

Birkbeck UNIVERSITY OF LONDON

Department of Economics, Mathematics and Statistics

Birkbeck Working Papers in Economics & Finance

BWPEF 1704

Causes and Effects of Negative Definite Covariance Matrices in Swamy Type Random Coefficient Models

Andrea Nocera
Birkbeck, University of London

June 2017

▪ Birkbeck, University of London ▪ Malet Street ▪ London ▪ WC1E 7HX ▪

Causes and Effects of Negative Definite Covariance Matrices in Swamy Type Random Coefficient Models

Andrea Nocera
Birkbeck, University of London*

June, 2017

Abstract

In this paper, we investigate the causes and the finite-sample consequences of negative definite covariance matrices in Swamy type random coefficient models. Monte Carlo experiments reveal that the negative definiteness problem is less severe when the degree of coefficient dispersion is substantial, and the precision of the regression disturbances is high. The sample size also plays a crucial role. We then demonstrate that relying on the asymptotic properties of a biased but consistent estimator of the random coefficient covariance may lead to poor inference.

JEL Classification: C12, C15, C23.
Keywords: Finite-sample inference, Monte Carlo analysis, negative definite covariance matrices, panel data, random coefficient models.

*Helpful comments by George Kapetanios, Zacharias Psaradakis, Ron Smith, and Takashi Yamagata are gratefully acknowledged. All remaining errors are mine.
E-mail address: a.nocera@mail.bbk.ac.uk

1 Introduction

For panel data studies with large N, the number of units, and small T, the time dimension, it is common to assume homogeneity of the slope coefficients. Individual-specific intercepts are the only source of heterogeneity. However, in many economic applications, it is more realistic to allow the response parameters to differ across cross-sectional units. As T increases, it is possible to test for equality of parameters, and the homogeneity hypothesis is very often rejected. Two popular methods which deal with coefficient heterogeneity are the Mean Group estimation, proposed by Pesaran and Smith (1995), and the Swamy (1970) random coefficient model. Both methods require estimating N time series separately. The latter models the regression coefficients as random variables with a certain probability distribution. To reduce the number of parameters to be estimated, it is assumed that the coefficients have constant means and variance-covariances.

Unfortunately, as in the error-component model, the estimator of the random coefficient covariance matrix is not necessarily nonnegative definite. This is often the case in empirical applications. Despite being a well acknowledged problem, its causes are not yet fully understood. In this paper, we disentangle the drivers of the problem by means of Monte Carlo simulations. Another contribution of this paper is to examine the finite-sample properties of Swamy's generalized least squares (GLS) estimator in terms of accuracy of inference, when a consistent but biased estimator of the random coefficient covariance is used to overcome the negative definiteness problem.

The Monte Carlo analysis confirms that the negative definiteness problem of this estimator increases with the variance of the regression time-varying disturbances, and it is negatively (and statistically significantly) correlated to the degree of coefficient heterogeneity. The probability of the estimator being negative definite goes much faster to zero following an increase in the level of coefficient dispersion rather than a raise in the precision of the regression disturbances. The problem is also more severe when T and/or N are small, partly due to the fact that the performances of individual OLS and the Mean Group estimators worsen in small samples. As expected, when T goes to infinity, the second term of the estimator goes to zero, and the problem of negative definiteness vanishes.

Whenever the unbiased estimator of the random coefficient covariance is negative definite, Swamy suggests eliminating a term to obtain an estimator which is nonnegative definite and is consistent when T tends to infinity. However, we show that the latter can be severely biased in small samples. We then investigate the finite-sample consequences for hypothesis tests. We find that the resulting estimated standard errors are very often upwards biased. In many cases, this bias can be substantial. This in turn leads to size distorted hypothesis tests, with exact sizes well below the nominal levels.

The remainder of the paper is organized as follows. Section 2 reviews the random coefficient model. Section 3 discusses the derivation of the Swamy estimator of the random coefficient covariance matrix. Monte Carlo experiments are implemented in Section 4, where we present the results from regressing the probability of the estimator being negative definite on a number of explanatory variables, and comment on the finite-sample performances of the estimator of interest for inference. The last section concludes.

2 The Random Coefficient Model

Consider the following linear regression model

$$y_i = X_i\beta_i + u_i, \qquad i = 1, .., N, \tag{1}$$

where $y_i = (y_{i1}, y_{i2}, ..., y_{iT})'$ is a $T \times 1$ vector of observations for the dependent variable, and X_i is a $T \times K$ matrix of strictly exogenous explanatory variables, including a vector of ones to allow for an intercept. The Swamy (1970) random coefficient model treats both intercept and slope coefficients

$$\beta_i = \beta + \delta_i \tag{2}$$

as random with common mean β. It is assumed that

$$E(\delta_i) = 0, \qquad E\left(\delta_i\delta_j'\right) = \begin{cases} \triangle & if \quad i = j \\ 0 & if \quad i \neq j \end{cases} \tag{3}$$

$$E(u_i) = 0, \qquad E\left(u_iu_j'\right) = \begin{cases} \sigma_i^2 I_T & if \quad i = j \\ 0 & if \quad i \neq j \end{cases} \tag{4}$$

Finally, β_i and u_j are independent for all i and j.

2.1 Estimation

Under the above assumptions, the best linear unbiased estimator of β is the generalized least squares (GLS) estimator

$$\begin{aligned} \hat{\beta}_{GLS} &= \left(\textstyle\sum_{i=1}^N X_i'V_i^{-1}X_i\right)^{-1}\left(\textstyle\sum_{i=1}^N X_i'V_i^{-1}y_i\right) \\ &= \textstyle\sum_{i=1}^N W_i\hat{\beta}_i, \end{aligned} \tag{5}$$

where

$$\begin{aligned} W_i &= \left\{\textstyle\sum_{i=1}^N \left[\triangle + \sigma_i^2(X_i'X_i)^{-1}\right]^{-1}\right\}^{-1}\left[\triangle + \sigma_i^2(X_i'X_i)^{-1}\right]^{-1}, \\ \hat{\beta}_i &= (X_i'X_i)^{-1}X_i'y_i, \end{aligned}$$

and $V_i = X_i \triangle X_i' + \sigma_i^2 I_T$. The GLS estimator is equivalent to the weighted average of the OLS estimates, with weights inversely proportional to their covariance matrices. The variance-covariance matrix of (5) is

$$var\left(\hat{\beta}_{GLS}\right) = \left(\sum_{i=1}^{N} X_i' V_i^{-1} X_i\right)^{-1}. \tag{6}$$

As noted by Swamy, if we assume normality of both u_i and β_i, it can be easily shown that the variance of the GLS estimator is equal to the Cramer-Rao lower bound. Therefore, (5) is a minimum variance estimator within the class of all unbiased estimators.

However, the GLS estimator for β is infeasible since it depends on the unknown variances σ_i^2 and $\triangle$. Swamy uses the OLS estimators, $\hat{\beta}_i$, and their residuals $\hat{u}_i = y_i - X_i\hat{\beta}_i$, to obtain unbiased estimators of σ_i^2 and $\triangle$,

$$\hat{\sigma}_i^2 = \frac{\hat{u}_i'\hat{u}_i}{T-K}, \tag{7}$$

$$\hat{\triangle} = \hat{\triangle}_1 - \hat{\triangle}_2, \tag{8}$$

where

$$\begin{aligned} \hat{\triangle}_1 &= \tfrac{1}{N-1}\textstyle\sum_{i=1}^{N}\left(\hat{\beta}_i - N^{-1}\sum_{i=1}^{N}\hat{\beta}_i\right)\left(\hat{\beta}_i - N^{-1}\sum_{i=1}^{N}\hat{\beta}_i\right)', \\ \hat{\triangle}_2 &= N^{-1}\textstyle\sum_{i=1}^{N}\hat{\sigma}_i^2 (X_i'X_i)^{-1}. \end{aligned} \tag{9}$$

The second term $\left(-\hat{\triangle}_2\right)$ is necessary for $\hat{\triangle}$ to be an unbiased estimator of $\triangle$. Unfortunately, as in the error-component model, the estimator (8) is not necessarily nonnegative definite. As a solution, Swamy suggested using $\hat{\triangle}_1$ as an estimator of $\triangle$. Although biased, this estimator is positive semi-definite and consistent when T tends to infinity. Note that as T gets large, the second term, $\hat{\triangle}_2$, converges in probability to zero.

3 Derivation of the Estimator of the Random Coefficient Covariance Matrix

In this section, we describe the derivation of (8) in some detail. We start by noting that the OLS estimator of β_i can be rewritten as

$$\begin{aligned} \hat{\beta}_i &= \beta_i + (X_i'X_i)^{-1}X_i'u_i \\ &= \beta + \delta_i + (X_i'X_i)^{-1}X_i'u_i. \end{aligned} \tag{10}$$

Its unconditional and conditional expectations are given by $E\left(\hat{\beta}_i\right)=\beta$, and $E\left(\hat{\beta}_i \mid \delta_i\right)=\beta_i$, respectively. Using equation (10), we can compute the variance of the OLS estimator:

$$\begin{aligned} var\left(\hat{\beta}_i\right) &= E\left(\hat{\beta}_i-\beta\right)\left(\hat{\beta}_i-\beta\right)' \\ &= E\left[\left(\beta_i-\beta\right)+\left(\hat{\beta}_i-\beta_i\right)\right]\left[\left(\beta_i-\beta\right)+\left(\hat{\beta}_i-\beta_i\right)\right]', \end{aligned} \tag{11}$$

where $(\beta_i-\beta)=\delta_i$, and $\left(\hat{\beta}_i-\beta_i\right)=(X_i'X_i)^{-1}X_i'u_i$. Using equations (3) and (4), and assuming that $E\left(u_i \mid X_i,\delta_i\right)=0$, we get

$$\begin{aligned} var\left(\hat{\beta}_i\right) = E\left(\hat{\beta}_i-\beta\right)\left(\hat{\beta}_i-\beta\right)' &= E\left(\beta_i-\beta\right)\left(\beta_i-\beta\right)'+E\left(\hat{\beta}_i-\beta_i\right)\left(\hat{\beta}_i-\beta_i\right)' \\ &= \triangle+\sigma_i^2(X_i'X_i)^{-1}. \end{aligned} \tag{12}$$

Equation (12) states that, for an unbiased estimator where $E\left(\hat{\beta}_i \mid \beta_i\right)=\beta_i$ and $E\left(\hat{\beta}_i\right)=\beta$, the variance of $\hat{\beta}_i$ around β is equal to the variance of β_i around β plus the variance of $\hat{\beta}_i$ around β_i.

The estimator of $\triangle$ given by (8), can be obtained by replacing $var\left(\hat{\beta}_i\right)$ with its sample analogue, and $\sigma_i^2(X_i'X_i)^{-1}$ with its estimator averaged across units.

From equation (12), it follows that

$$\triangle=E\left(\beta_i-\beta\right)\left(\beta_i-\beta\right)'=\triangle_1-\triangle_2, \tag{13}$$

where

$$\begin{aligned} \triangle_1 &= E\left(\hat{\beta}_i-\beta\right)\left(\hat{\beta}_i-\beta\right)', \\ \triangle_2 &= E_{\hat{\beta}_i|\beta_i}\left[\hat{\beta}_i-E\left(\hat{\beta}_i \mid \beta_i\right)\right]\left[\hat{\beta}_i-E\left(\hat{\beta}_i \mid \beta_i\right)\right]'. \end{aligned} \tag{14}$$

It can be noted that $\triangle$ is positive semi-definite by definition. Indeed,

$$\left\{\begin{aligned} \hat{\beta}_i-\beta &= \delta_i+(X_i'X_i)^{-1}X_i'u_i \\ \hat{\beta}_i-\beta_i &= (X_i'X_i)^{-1}X_i'u_i \end{aligned}\right. \Longrightarrow E\left(\hat{\beta}_i-\beta\right)\left(\hat{\beta}_i-\beta\right)' \geq E\left(\hat{\beta}_i-\beta_i\right)\left(\hat{\beta}_i-\beta_i\right)', \tag{15}$$

where $\beta_i=E\left(\hat{\beta}_i \mid \beta_i\right)$, and the inequality sign denotes matrix inequalities. The equality would hold only if $\delta_i=0$, $\forall i$, which means that the coefficients do not vary across units, i.e. $E\left(\delta_i\delta_i'\right)=0$, for all i.

It can also be noted that (12) satisfies the law of total variance since

$$\begin{aligned} var\left(\hat{\beta}_i\right) &= var\left[E\left(\hat{\beta}_i \mid \beta_i\right)\right] + E\left[var\left(\hat{\beta}_i \mid \beta_i\right)\right] \\ &= var\left(\beta_i\right) + E\left[var\left(\hat{\beta}_i \mid \beta_i\right)\right], \end{aligned}$$

where $var\left(\beta_i\right) = \triangle$, and $var\left(\hat{\beta}_i \mid \beta_i\right) = \sigma_i^2(X_i'X_i)^{-1}$. This implies that $var\left(\hat{\beta}_i\right) \geq \triangle$, and $var\left(\hat{\beta}_i\right) \geq \sigma_i^2(X_i'X_i)^{-1}$, which corroborates (15).[1]

3.1 Nonspherical Errors

Equation (12) has been derived under the assumption that $var\left(u_i\right) = \sigma_i^2 I_T$. If $var\left(u_i\right) = \Omega_i$, where Ω_i is a symmetric and positive definite $T \times T$ matrix, then

$$V_i = X_i \triangle X_i' + \Omega_i,$$

and

$$\begin{aligned} E\left(\hat{\beta}_i - \beta_i\right)\left(\hat{\beta}_i - \beta_i\right)' &= E\left((X_i'X_i)^{-1}X_i'u_iu_i'X_i(X_i'X_i)^{-1}\right) \\ &= \left(X_i'X_i\right)^{-1}\left(X_i'\Omega_iX_i\right)\left(X_i'X_i\right)^{-1}. \end{aligned}$$

Equation (11) becomes

$$var\left(\hat{\beta}_i\right) = \triangle + \left(X_i'X_i\right)^{-1}\left(X_i'\Omega_iX_i\right)\left(X_i'X_i\right)^{-1}. \tag{16}$$

Therefore, an unbiased estimator of $\triangle$ is

$$\hat{\triangle} = \hat{\triangle}_1 - \hat{\triangle}_3, \tag{17}$$

where $\hat{\triangle}_1$ is defined in (9), and

$$\hat{\triangle}_3 = \frac{1}{N}\sum_{i=1}^{N}\left(X_i'X_i\right)^{-1}\left(X_i'\hat{\Omega}_iX_i\right)\left(X_i'X_i\right)^{-1}. \tag{18}$$

In many cases, $\hat{\triangle}_3 \geq \hat{\triangle}_2$, which may exacerbate the negative definiteness problem of $\hat{\triangle}$. When the elements of u_i are negatively autoccorelated, and the K regressors in x_{it} are positively autocorrelated, Goldeberger (1964, pp. 238-42) showed that the diagonal elements of $\hat{\triangle}_3$ can be smaller than the corresponding diagonal elements of $\hat{\triangle}_2$.

[1] Henceforth, we use inequality signs to denote matrix inequalities.

Alternatively, as shown in Appendix A.1, one can estimate each time series by applying Aitken's GLS, yielding

$$\tilde{\beta}_i = (X_i'\hat{\Omega}_i^{-1}X_i)^{-1}X_i'\hat{\Omega}_i^{-1}y_i,$$

$$\tilde{\sigma}_i^2 = \frac{\tilde{u}_i'\tilde{u}_i}{T-K},$$

where $\tilde{u}_i$ are the GLS residuals. In such case, the estimator of $\triangle$ becomes

$$\hat{\triangle} = \hat{\triangle}_4 - \hat{\triangle}_5, \tag{19}$$

where

$$\begin{aligned} \hat{\triangle}_4 &= \tfrac{1}{N-1}\textstyle\sum_{i=1}^N \left(\tilde{\beta}_i - N^{-1}\sum_{i=1}^N \tilde{\beta}_i\right)\left(\tilde{\beta}_i - N^{-1}\sum_{i=1}^N \tilde{\beta}_i\right)', \\ \hat{\triangle}_5 &= \tfrac{1}{N}\textstyle\sum_{i=1}^N \tilde{\sigma}_i^2 \left(X_i'\hat{\Omega}_iX_i\right)^{-1}, \end{aligned} \tag{20}$$

It is reasonable to expect $\hat{\triangle}_5$ to be smaller than $\hat{\triangle}_3$.

One may suspect that taking serial correlation into account, and using (19) as an estimator of $\hat{\triangle}$ reduces the probability of $\hat{\triangle}$ being negative definite.[2] For instance, Swamy (1971) indicates mispecification of either the model or the underlying assumptions as possible reasons of the negative definiteness problem. Nevertheless, as shown in the Monte Carlo analysis, $\hat{\triangle}$ can be often negative definite even though the true disturbances are not correlated over time and the model is correctly specified, suggesting that the causes of the problem lie elsewhere.

4 Monte Carlo Analysis

Given that $\triangle$ is positive semi-definite by construction, why is (8) often negative semi-definite? What goes wrong when replacing the true components with their analogue estimates? In other words, why is $\hat{\triangle}_1$, the estimator of $E\left(\hat{\beta}_i - \beta\right)\left(\hat{\beta}_i - \beta\right)'$, often less (in a matrix sense) than $\hat{\triangle}_2$, the estimator of $E\left(\hat{\beta}_i - \beta_i\right)\left(\hat{\beta}_i - \beta_i\right)'$? We address this question by performing a Monte Carlo analysis. We then evaluate the direct consequences of relying on the asymptotic properties of a biased but consistent estimator of the random coefficient covariance for hypothesis tests.

[2]It should be noted that although $\hat{\triangle}_5$ might be smaller than $\hat{\triangle}_3$, $\hat{\triangle}_1$ has to be replaced by $\hat{\triangle}_4$ in estimating $\hat{\triangle}$. As for (8), there is no guarantee that (19) is nonnegative definite.

4.1 The Data Generating Process

The data generating process used to simulate the data is given by

$$\begin{aligned} y_{it} &= c_i + x_{it}\beta_i + \varepsilon_{it}, \\ x_{it} &= c_{x,i}(1-\rho) + \rho x_{it-1} + u_{it}, \end{aligned} \tag{21}$$

where

$$\begin{aligned} \varepsilon_{it} &\sim i.i.d.N(0,\sigma_i^2), \\ u_{it} &\sim i.i.d.N(0,1), \\ c_{x,i} &\sim i.i.d.N(1,1). \end{aligned}$$

We set $\rho = 0.6$, and $x_{i0} = 0, \forall i$. Once generated, the x_{it}'s are taken as fixed across different replications.[3] The variances of the time-varying disturbances are generated according to:

$$\begin{array}{ccccc} (i) & \sigma_i^2 & \sim & unif\,[0.1, 0.9]\,, \\ (ii) & \sigma_i^2 & \sim & unif\,[0.5, 1.5]\,, \\ (iii) & \sigma_i^2 & \sim & unif\,[1, 3]\,, \\ (iv) & \sigma_i^2 & \sim & unif\,[3, 5]\,, \end{array}$$

such that $E(\sigma_i^2) \in \{0.5, 1, 2, 4\}$. To allow for the presence of outliers, we also consider the following case

$$(v) \quad \sigma_i^2 \quad \sim \quad \varphi \cdot unif\,[0.5, 1.5] + (1-\varphi) \cdot unif\,[4, 6]\,,$$

where φ is binary variable whose distribution is Bernoulli:

$$\varphi = \begin{cases} 1 & p = 0.75 \\ \\ 0 & (1-p). \end{cases}$$

In the latter case, $E(\sigma_i^2) = 2$ as in case (iii) although the variance of most of the units varies between 0.5 and 1.5. In all cases, the σ_i^2 's are sorted so that $\sigma_i^2 > \sigma_j^2$ if $\bar{x}_i > \bar{x}_j$, where $\bar{x}_i = T^{-1}\sum_{t=1}^{T} x_{it}$. The coefficients differ randomly across units according to

$$\begin{aligned} c_i &= c + \sigma_c \gamma_{1i}, \\ \beta_i &= \beta + \sigma_\beta \gamma_{2i}, \end{aligned}$$

where $\gamma_{ji} \sim i.i.d.N(0,1)$, for $j = 1, 2$. We consider the following options:

[3]To minimize the effect of initial observations we discard the first 100 observations.

Option	1	2	3	4	5	6	7
c	0	0	0	0.5	0.5	0.5	1
β	0.1	0.5	1	0.1	0.5	1	1

For each option, we draw the random effects, γ_{ji}, from a Normal distribution with different degrees of coefficient heterogeneity (from low (1) to high (6)):[4]

Degree of Heterogeneity	1	2	3	4	5
σ_c	0.05	0.1	0.3	0.5	1
σ_β	0.05	0.1	0.3	0.5	1

We generate $G = (n_O \cdot n_H) \cdot n_V = (7 \cdot 5) \cdot 5 = 175$ clusters, where n_O, n_H, and n_V denote the number of options, the number of coefficient heterogeneity cases, and the different specifications for σ_i^2, respectively. Each cluster is of size $S = (n_T \cdot n_N) = (6 \cdot 4) = 24$, where each unit in the cluster consists of the pair (T_j, N_l), with $T_j \in \{10, 20, 30, 50, 70, 140\}$, and $N_l \in \{10, 30, 50, 140\}$. In total, we run $M = (n_T \cdot n_N)(n_O \cdot n_H \cdot n_V) = 24 \cdot 175 = 4200$ different data generating processes (DGP). Within each DGP we run $H = 1500$ iterations.[5]

Degree of Coefficient Heterogeneity. The choice of σ_c and σ_β is in line with Trapani and Urga (2009), and Boyd and Smith (2002). The former review some empirical works which use heterogenous estimators and derive a measure to determine the level of coefficient heterogeneity (the standard deviation of the random coefficients). They find that the levels of heterogeneity obtained using the datasets of Baltagi, Griffin and Xiong (2000) and Baltagi, Bresson, Griffin and Pirotte (2003) are equal to 0.176 and 0.183 respectively. Higher levels are found in Baltagi and Griffin (1997) and Brucker and Siliverstovs (2006), where the degree of heterogeneity is equal to 0.323 and 0.428 respectively.

Boyd and Smith (2002) review some econometric issues in estimating models of the transmission mechanism of monetary policy, for 57 developing countries, where $T = 31$. They find a high degree of dispersion of the estimates across countries.[6] For instance, in an inflation

[4]It should be noted that when generating 1000 observations from $\beta_i \sim N(0.5, 1)$, the range of values that β_i assumed was -3 to 3.4. This is a very high level of dispersion, which we consider for theoretical reasons. If the degree of heterogeneity were so high in real applications, it might be difficult to reconcile the estimates with economic theory.

[5]The time required to run the 1500 iterations is approximately 5 to 30 seconds depending on the sample size. The time necessary to estimate each option is approximately 2 hours, which makes the results replicable.

[6]After estimating the regression coefficients, β_i, for each unit, Boyd and Smith compute the number of standard deviations from the mean as $Z(\beta) = \left(\hat{\beta}_i - \bar{\beta}\right) / s\left(\hat{\beta}_i\right)$, where $\bar{\beta} = N^{-1} \sum_{i=1}^{N} \hat{\beta}_i$, and $s^2\left(\hat{\beta}_i\right) = (N-1)^{-1} \sum_{i=1}^{N} \left(\hat{\beta}_i - \bar{\beta}\right)^2$.

persistence equation, the average coefficient on the first lag of inflation is 0.57 with a standard deviation of 0.30. In a static Purchasing Power Parity equation of log spot on log price differential, the mean is 1.13 and the standard deviation of the estimates is 0.52.

4.2 Descriptive Statistics

Table 1 reports the Monte Carlo estimates of $\varrho = Pr\left(\hat{\triangle} < 0\right)$, the probability that the estimator of $\triangle$ (defined in (8), and averaged across the 7 different options) is negative definite, across different sample sizes and different combinations of coefficient and data dispersions, σ_β and $E\left(\sigma_i\right)$ respectively.

Table 1: The probability of $\hat{\triangle}$ being negative definite

		$E\left(\sigma_i^2\right) = 2$ (iii)					$E\left(\sigma_i^2\right) = 2$ (v)					$E\left(\sigma_i^2\right) = 1$				
T	$N \setminus \sigma_\beta$	0.05	0.1	0.3	0.5	1	0.05	0.1	0.3	0.5	1	0.05	0.1	0.3	0.5	1
10	10	88	86	73	57	33	88	86	75	57	41	87	84	66	44	11
	30	**84**	**83**	**63**	**43**	**8**	**85**	**83**	**63**	**54**	**19**	**84**	**80**	**51**	**27**	**1**
	50	84	80	61	36	5	84	80	58	45	17	82	77	43	15	0
	140	81	77	46	24	0	81	76	49	37	3	79	71	32	4	0
20	10	88	85	64	37	7	89	85	61	41	9	87	80	45	18	1
	30	**84**	**79**	**46**	**13**	**0**	**84**	**75**	**51**	**25**	**2**	**80**	**68**	**20**	**2**	**0**
	50	82	72	33	7	0	83	74	44	21	0	79	64	11	0	0
	140	78	64	15	1	0	79	63	33	5	0	73	52	2	0	0
30	10	88	82	49	20	2	89	81	52	33	4	86	77	33	11	0
	30	**84**	**71**	**24**	**2**	**0**	**85**	**70**	**39**	**11**	**0**	**79**	**61**	**6**	**0**	**0**
	50	81	66	18	1	0	79	67	32	5	0	74	54	2	0	0
	140	75	54	3	0	0	75	56	14	0	0	66	39	0	0	0
50	10	87	78	34	9	1	88	78	40	7	1	83	68	17	3	0
	30	**80**	**62**	**10**	**0**	**0**	**80**	**65**	**20**	**1**	**0**	**74**	**49**	**1**	**0**	**0**
	50	75	56	3	0	0	76	57	9	0	0	66	42	0	0	0
	140	67	41	0	0	0	66	49	2	0	0	56	25	0	0	0
140	10	80	62	6	0	0	80	66	12	1	0	73	49	2	0	0
	30	**66**	**39**	**0**	**0**	**0**	**67**	**48**	**1**	**0**	**0**	**54**	**18**	**0**	**0**	**0**
	50	60	28	0	0	0	61	39	0	0	0	46	10	0	0	0
	140	46	11	0	0	0	52	27	0	0	0	31	1	0	0	0

The probability (in percentage) of the estimator of $\triangle$ being negative definite (averaged across the 7 options) across the time dimension (T), the cross-section dimension (N), different degrees of coefficient heterogeneity (σ_β), and the mean of the variance of the time-varying regression disturbances ($E(\sigma_i^2)$), for $i = 1, .., N$. The results shown in columns (iii) and (v) differ as in the former $\sigma_i^2 \sim unif\,[1, 3]$. In the latter, $\sigma_i^2 \sim \varphi \cdot unif\,[0.5, 1.5] + (1 - \varphi) \cdot unif\,[4, 6]$.

A few important facts emerge from this simple descriptive analysis:

1. The probability ϱ is a decreasing function of both T and N. However, when T and σ_β are moderate, the probability of $\hat{\triangle}$ being negative definite can still be high even when N is as large as 140.

2. ϱ can be quite high when σ_β is small or moderate. If $\sigma_\beta = 0.05$, the value of ϱ can be substantial even when $T = 140$ and N is also large. On the contrary, if $\sigma_\beta = 1$, ϱ is almost always equal to zero as soon as T is larger than 20.

3. The variance of the time-varying disturbances also plays an important role. Indeed, for a given degree of coefficient heterogeneity, as σ_i^2 increases, the second term of (8) raises. Consequently, the probability that the estimator of the random coefficient covariance matrix is negative definite increases.

4. Whether ϱ is large or small depends on the value of σ_β relative to the σ_i's, the standard deviations of the time-varying regression disturbances. This means that even though σ_β is high, ϱ can be still far from zero if $E\left(\sigma_i^2\right)$ is very large.

4.3 Regression Analysis

To corroborates the findings of the theoretical analysis and the insights emerged in the descriptive analysis, we run the following cross-section regression:

$$y_m = \alpha + z_m'\theta + u_m, \qquad m = 1, .., M,$$

where $M = 4200$. The depedent variable y_m measures the probability of $\hat{\triangle}$ being negative definite within each DGP, and it is computed as

$$y_m = Pr\left(\hat{\triangle} < 0\right) = \frac{\sum_{h=1}^{H} \rrbracket\left(\hat{\triangle}^{(h)} < 0\right)}{H},$$

where $\rrbracket$ is a binary indicator that takes the value 1 if $\hat{\triangle}^{(h)} < 0$ (in a matrix sense) and 0 otherwise. The vector z_m may include the following explanatory variables:

- the time dimension, T, and the number of units, N,
- the values of the intercept (c) and slope parameter (β) in (21),
- the degree of coefficient heterogeneity, $\sigma_c = \sigma_\beta$,
- the average standard deviation of the regression disturbances, $\bar{\sigma} = N^{-1} \sum_{i=1}^{N} \sigma_i$,

- a measure of the signal-to-noise ratio, $\sigma_\beta/\bar{\sigma}$,
- the bias of the Mean Group estimator of $\psi = (c, \beta)'$,
- the cross-section averages of the absolute value of the biases of the OLS estimators:

$$\frac{1}{N}\sum_{i=1}^{N}\left|\left(\frac{1}{H}\sum_{h=1}^{H}\hat{\psi}_i^{(h)}\right)-\psi\right|,$$

- the trace of the root mean square errors (RMSE) of the Mean Group estimator,
- the trace of the RMSE of the OLS estimators, averaged across units:

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\sqrt{\frac{1}{H}\sum_{h=1}^{H}\left(\hat{\psi}_i^{(h)}-\psi_i^{(h)}\right)\left(\hat{\psi}_i^{(h)}-\psi_i^{(h)}\right)'}\right\}.$$

We estimate the model by OLS. Results are shown in Table 2. In parenthesis, we report the t-tests computed using White (1980) heteroskedasticity-robust standard errors.[7]

Main Findings. In the simplest specification (1), we regress our dependent variables on a constant, the time dimension (T), the number of units (N), the degree of coefficient heterogeneity (σ_β), and the average of the time-varying regression disturbances' standard deviations ($\bar{\sigma}$). We then include the value of c and β used in equation (21) to simulate the data. As expected, the constant, which is approximatively equal to 70%, is statistically significant. One standard deviation increase of σ_β statistically significantly reduces the probability of $\hat{\triangle}$ being negative definite (ϱ) of around 70%. The conditional variability of the data is also a significant predictor: one standard deviation increase of $\bar{\sigma}$ is associated with a statistically significant increase in the dependent variable of 5%. At the same time, an one unit increase in T and N causes a 0.2% and 0.1% decrease of ϱ, respectively. On the contrary, the coefficients associated with the value of the constant and intercept parameters (c and β) are not statistically significant. These findings are consistent across all other specifications.

[7]When calculating the robust standard errors, we make the adjustment for degrees of freedom suggested by MacKinnon and White (1985).

Table 2: The drivers of the random coefficient covariance's negative definiteness problem

$Pr\left(\hat{\triangle}<0\right)$	(1)	(2)	(3)	(4)	(5)	(6)
constant	**0.696**	**0.695**	**0.744**	**0.562**	**0.521**	**0.536**
	(82.77)	(71.27)	(77.33)	(58.52)	(46.54)	45.031
T	**-0.002**	**-0.002**	**-0.002**	**-0.001**	**-0.001**	**-0.001**
	(-32.04)	(-32.03)	(-28.99)	(-12.85)	(-9.38)	-11.596
N	**-0.001**	**-0.001**	**-0.001**	**-0.001**	**-0.001**	**-0.001**
	(-21.52)	(-21.52)	(-19.09)	(-23.73)	(-8.41)	-10.312
σ_β	**-0.697**	**-0.697**		**-1.160**	**-0.802**	**-1.218**
	(-91.69)	(-91.66)		(-58.12)	(-71.60)	-45.356
$\bar{\sigma}$	**0.052**	**0.052**		**0.018**	**0.015**	**0.020**
	(21.59)	(21.60)		(7.31)	(5.31)	7.318
$\sigma_\beta/\bar{\sigma}$			**-0.639**			
			(-49.98)			
c		0.003	0.002			0.004
		(0.31)	(0.18)			0.477
β		0.001	0.001			-0.002
		(0.11)	(0.152)			-0.333
$bias\left(\hat{c}_{mg}\right)$				-0.198		-0.083
				(-0.32)		-0.132
$bias\left(\hat{\beta}_{mg}\right)$				0.239		0.347
				(0.30)		0.449
$Av\left(\left\|bias\left(\hat{c}_{i,ols}\right)\right\|\right)$				**13.099**		**12.346**
				(17.26)		12.215
$Av\left(\left\|bias\left(\hat{\beta}_{i,ols}\right)\right\|\right)$				**14.246**		**13.366**
				(12.09)		11.058
$RMSE\left(\hat{\psi}_{MG}\right)$					**0.373**	**0.301**
					(14.78)	10.777
$RMSE\left(\hat{\psi}_{i,ols}\right)$					**0.163**	**-0.044**
					(16.90)	-2.863
R^2	0.675	0.675	0.576	0.728	0.713	0.734
Theil Adj. R^2	0.675	0.675	0.576	0.728	0.713	0.733

We regress the probability of $\hat{\triangle}$ being negative definite on a number of explanatory variables. The values of the OLS estimators and their corresponding t-ratios (in parentheses) are reported. We use White (1980) heteroskedasticity-robust standard errors with the adjustment for degrees of freedom suggested by MacKinnon and White (1985). Bold values denotes statistical significance at 5% level or lower.

In a third regression (3), we replace σ_β and $\bar{\sigma}$ with a measure of the signal-to-noise ratio $(\sigma_\beta/\bar{\sigma})$.[8] An one standard deviation increase of the latter statistically significantly decreases the probability of $\hat{\triangle}$ being negative definite by 64%. The R-squared is smaller in the third specification, suggesting that including both σ_β and $\bar{\sigma}$ separately improves the goodness of fits.

Given that $\hat{\triangle}$, described in equation (8), is a plug-in estimator, we also test whether the finite sample performances (in terms of bias and RMSE) of both the Mean Group estimator of c and β, and the OLS estimators of the unit-specific regression coefficients affect the probability of $\hat{\triangle}$ being negative definite. The regression analyses (4) to (6) corroborate this hypothesis. For instance, a 1% increase in the cross-section averages of the absolute value of the biases of the OLS estimates raises ϱ of around 12 to 14%.

4.4 Finite-Sample Consequences

As shown in Table 1, the unbiased estimator of the random coefficient covariance matrix defined in equation (8), is likely to be negative definite in many circumstances. This is often the case in many empirical applications. To overcome the problem, Swamy (1971) suggests replacing this estimator by $\hat{\triangle}_1$, defined in equation (9). The latter is nonnegative definite and is consistent when T tends to infinity. However, as reported in Table 3, it can be severely biased in small samples.

Therefore, it is important to assess the finite-sample consequences of using $\hat{\triangle}_1$ as an estimator of $\triangle$. The aim of this subsection is to provide some evidence on whether it is appropriate to rely on the asymptotic properties of this estimator as the basis for inference in finite samples. Without loss of generality, we focus on the results obtained from Option 2, where $(c, \beta) = (0, 0.5)$. We only show results obtained when $E\left(\sigma_i^2\right) = 1$, for various degrees of coefficient heterogeneity. The consequences of using $\hat{\triangle}_1$ as an estimator of $\triangle$ are even more severe when $E\left(\sigma_i^2\right)$ increases.[9] Further analyses are available in an online Appendix.

[8] We have also considered other measures of signal-to-noise ratio: $N^{-1}\sum_{i=1}^{N}\left(\sigma_\beta^2/\sigma_{\varepsilon_i}^2\right)$, $N^{-1}\sum_{i=1}^{N}\left(\sigma_\beta/\sigma_{\varepsilon_i}\right)$, and $\left(\sigma_\beta^2/\bar{\sigma}^2\right)$, where $\bar{\sigma}^2 = N^{-1}\sum_{i=1}^{N}\sigma_{\varepsilon_i}^2$. They yield very similar result. Therefore, we only report results obtained using $(\sigma_\beta/\bar{\sigma})$, with $\bar{\sigma} = N^{-1}\sum_{i=1}^{N}\sigma_{\varepsilon_i}$ as the corresponding regression coefficient has larger economic value and it is associated with a larger t-ratio. Both the R^2 and the Theil's adjusted R^2 are also relatively larger in the latter case.

[9] Case (v) is particularly interesting. Even though the variance of most of the units varies between 0.5 and 1.5, as in case (ii), the presence of some outliers, such that $E\left(\sigma_i^2\right) = 2$, considerably worsen the accuracy of inference.

Table 3: Bias and root mean square errors of $\hat{\triangle}_1$

		$\sigma_\beta = 0.05$				$\sigma_\beta = 0.1$				$\sigma_\beta = 0.3$				$\sigma_\beta = 0.5$			
	$T \backslash N$	10	30	50	140	10	30	50	140	10	30	50	140	10	30	50	140
$bias\{\hat{\sigma}_c\}$	10	0.80	0.81	0.72	0.65	0.70	0.84	0.61	0.67	0.40	0.44	0.45	0.53	0.23	0.51	0.37	0.39
	20	0.39	0.34	0.41	0.38	0.32	0.35	0.32	0.34	0.17	0.19	0.21	0.19	0.13	0.15	0.13	0.14
	30	0.28	0.25	0.30	0.28	0.22	0.27	0.23	0.24	0.14	0.12	0.13	0.15	0.10	0.11	0.08	0.10
	50	0.21	0.22	0.20	0.19	0.18	0.18	0.15	0.17	0.06	0.10	0.08	0.09	0.06	0.05	0.05	0.05
	70	0.17	0.18	0.15	0.16	0.12	0.12	0.12	0.13	0.05	0.07	0.06	0.06	0.02	0.04	0.03	0.04
	140	0.10	0.10	0.11	0.10	0.08	0.07	0.07	0.07	0.01	0.03	0.03	0.03	0.00	0.01	0.01	0.02
$bias\{\hat{\sigma}_\beta\}$	10	0.39	0.33	0.31	0.32	0.35	0.33	0.22	0.27	0.14	0.15	0.18	0.19	0.06	0.13	0.13	0.11
	20	0.20	0.15	0.18	0.17	0.14	0.13	0.15	0.14	0.08	0.07	0.07	0.07	0.02	0.04	0.04	0.04
	30	0.12	0.13	0.12	0.12	0.08	0.10	0.09	0.09	0.03	0.04	0.04	0.04	0.03	0.03	0.02	0.03
	50	0.08	0.08	0.08	0.08	0.05	0.05	0.06	0.06	0.02	0.02	0.02	0.02	0.00	0.01	0.01	0.01
	70	0.07	0.06	0.06	0.06	0.04	0.04	0.04	0.04	0.01	0.01	0.01	0.02	0.00	0.00	0.01	0.01
	140	0.04	0.04	0.04	0.04	0.02	0.02	0.02	0.02	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00
$RMSE\{\hat{\sigma}_c\}$	10	0.84	0.83	0.74	0.66	0.74	0.87	0.62	0.67	0.45	0.46	0.46	0.54	0.29	0.54	0.39	0.40
	20	0.41	0.35	0.42	0.39	0.35	0.36	0.33	0.35	0.21	0.20	0.22	0.20	0.20	0.19	0.15	0.15
	30	0.30	0.26	0.30	0.28	0.24	0.28	0.24	0.24	0.18	0.13	0.14	0.15	0.18	0.14	0.10	0.11
	50	0.23	0.23	0.20	0.19	0.20	0.18	0.16	0.17	0.10	0.12	0.09	0.10	0.15	0.09	0.08	0.06
	70	0.18	0.19	0.15	0.16	0.13	0.13	0.12	0.13	0.10	0.09	0.07	0.06	0.13	0.08	0.06	0.05
	140	0.11	0.11	0.11	0.10	0.09	0.08	0.07	0.07	0.08	0.05	0.05	0.04	0.12	0.07	0.05	0.04
$RMSE\{\hat{\sigma}_\beta\}$	10	0.41	0.34	0.31	0.32	0.37	0.34	0.23	0.28	0.18	0.16	0.19	0.19	0.15	0.16	0.14	0.12
	20	0.21	0.15	0.18	0.17	0.15	0.13	0.15	0.14	0.12	0.09	0.08	0.07	0.13	0.08	0.07	0.05
	30	0.12	0.13	0.12	0.12	0.09	0.11	0.09	0.09	0.09	0.06	0.05	0.05	0.13	0.07	0.06	0.04
	50	0.08	0.08	0.08	0.08	0.06	0.06	0.06	0.06	0.08	0.05	0.04	0.03	0.12	0.07	0.05	0.03
	70	0.07	0.06	0.06	0.06	0.05	0.05	0.04	0.04	0.07	0.04	0.03	0.02	0.12	0.07	0.05	0.03
	140	0.04	0.04	0.04	0.04	0.04	0.02	0.02	0.02	0.07	0.04	0.03	0.02	0.12	0.07	0.05	0.03

The bias and root mean square errors (RMSE) of the square root of the diagonal elements of $\hat{\triangle}_1$, when $E\left(\sigma_i^2\right) = 1$ and $(c, \beta) = (0, 0.5)$ (Option 2), for various degree of coefficient heterogeneity ($\sigma_\beta = \sigma_c$), across different time (T) and cross-section dimensions (N).

Notation. Hereafter, we use the following notation to avoid repetition. We let $\psi_0 = (c, \beta)' = (0, 0.5)'$ be the true vector of average effects. The true random coefficient covariance matrix, $\triangle$, is diagonal, where σ_c^2 and σ_β^2 are the $(1,1)$ and $(2,2)$ entries, respectively. We let

$$\hat{\psi}_{GLS} = \left(\sum_{i=1}^{N} X_i' V_i^{-1} X_i \right)^{-1} \left(\sum_{i=1}^{N} X_i' V_i^{-1} y_i \right), \tag{22}$$

and

$$\Phi = var\left(\hat{\psi}_{GLS}\right) = \left(\sum_{i=1}^{N} X_i' V_i^{-1} X_i \right)^{-1}, \tag{23}$$

where $V_i = X_i \triangle X_i' + \sigma_i^2 I_T$, be the infeasible GLS estimator of ψ, and the infeasible covariance matrix of $\hat{\psi}_{GLS}$, respectively. The feasible GLS estimator, $\hat{\psi}_{FGLS}$, and an estimator of Φ, denoted $\hat{\Phi}$, are obtained by replacing σ_i^2 and $\triangle$ by $\hat{\sigma}_i^2$ and $\hat{\triangle}_1$, as defined in (7) and (9), respectively.

4.4.1 Accuracy of Estimated Standard Errors

To examine the consequences of overestimating the true random coefficient variances when testing hypotheses, we consider the ratio of the estimated standard errors (of the average effects) to the infeasible standard errors, obtained by taking the square root of the diagonal elements of $\hat{\Phi}$ and Φ respectively. Another measure of interest for inference is the accuracy of the estimated standard errors as approximations to the correct sampling standard deviation of the estimator of ψ.[10] These ratios should ideally be equal to one.

Table 4: Accuracy of estimated standard errors

		$\sigma_\beta = 0.05$				$\sigma_\beta = 0.1$				$\sigma_\beta = 0.3$				$\sigma_\beta = 0.5$			
	$T \backslash N$	10	30	50	140	10	30	50	140	10	30	50	140	10	30	50	140
$Accuracy\{se(\hat{c})\}$	10	1.69	1.77	1.84	1.88	1.67	2.17	1.80	1.86	1.47	1.53	1.62	1.69	1.22	1.51	1.46	1.41
	20	1.61	1.65	1.87	1.83	1.53	1.72	1.72	1.75	1.31	1.30	1.33	1.40	1.16	1.21	1.20	1.16
	30	**1.60**	**1.62**	**1.78**	**1.72**	**1.58**	**1.78**	**1.63**	**1.67**	**1.32**	**1.22**	**1.28**	**1.33**	**1.18**	**1.16**	**1.11**	**1.10**
	50	1.54	1.84	1.71	1.72	1.51	1.57	1.58	1.59	1.16	1.20	1.20	1.21	1.11	1.13	1.10	1.05
	70	1.65	1.65	1.61	1.66	1.42	1.46	1.53	1.51	1.12	1.18	1.18	1.17	1.02	1.08	1.07	1.09
	140	1.59	1.57	1.71	1.56	1.38	1.35	1.37	1.38	1.04	1.10	1.07	1.11	0.96	1.00	1.00	1.05
$Accuracy\{se(\hat{\beta})\}$	10	1.66	1.71	1.63	1.71	1.67	1.75	1.62	1.60	1.26	1.30	1.33	1.33	1.10	1.23	1.20	1.19
	20	1.53	1.53	1.62	1.55	1.41	1.49	1.52	1.54	1.17	1.17	1.16	1.15	1.02	1.06	1.10	1.06
	30	**1.50**	**1.53**	**1.57**	**1.53**	**1.36**	**1.46**	**1.40**	**1.45**	**1.09**	**1.12**	**1.13**	**1.11**	**1.06**	**1.06**	**1.05**	**1.03**
	50	1.40	1.54	1.53	1.47	1.23	1.35	1.35	1.30	1.06	1.02	1.09	1.09	1.01	1.02	1.03	1.03
	70	1.45	1.48	1.45	1.49	1.24	1.26	1.27	1.27	1.05	1.03	1.00	1.05	0.99	1.00	1.00	1.04
	140	1.39	1.38	1.37	1.38	1.19	1.16	1.12	1.16	1.02	1.01	1.01	1.00	0.96	1.01	0.99	1.01
$Ratio\{se(\hat{c})\}$	10	2.51	2.60	2.48	2.46	2.52	3.01	2.36	2.55	1.60	1.76	1.78	1.90	1.26	1.63	1.47	1.49
	20	2.12	2.03	2.29	2.20	1.70	2.03	1.95	2.01	1.34	1.37	1.41	1.40	1.18	1.23	1.19	1.21
	30	**1.87**	**1.82**	**2.05**	**1.99**	**1.77**	**1.99**	**1.82**	**1.82**	**1.29**	**1.26**	**1.28**	**1.33**	**1.15**	**1.17**	**1.13**	**1.16**
	50	1.74	2.13	1.96	1.87	1.72	1.74	1.65	1.70	1.15	1.24	1.20	1.23	1.10	1.09	1.09	1.09
	70	1.90	1.91	1.87	1.86	1.52	1.57	1.53	1.60	1.13	1.18	1.16	1.15	1.04	1.07	1.05	1.07
	140	1.74	1.73	1.78	1.72	1.43	1.40	1.41	1.41	1.04	1.08	1.09	1.09	1.00	1.02	1.03	1.04
$Ratio\{se(\hat{\beta})\}$	10	2.39	2.26	2.09	2.16	2.28	2.31	1.94	2.04	1.30	1.35	1.41	1.45	1.11	1.23	1.21	1.18
	20	1.91	1.78	1.99	1.87	1.50	1.63	1.62	1.66	1.20	1.19	1.18	1.18	1.04	1.07	1.07	1.08
	30	**1.60**	**1.65**	**1.75**	**1.72**	**1.49**	**1.57**	**1.51**	**1.50**	**1.10**	**1.11**	**1.11**	**1.13**	**1.05**	**1.05**	**1.05**	**1.05**
	50	1.52	1.63	1.65	1.61	1.33	1.37	1.37	1.40	1.07	1.06	1.07	1.07	1.01	1.02	1.02	1.03
	70	1.60	1.57	1.58	1.57	1.27	1.31	1.29	1.30	1.04	1.05	1.05	1.05	0.99	1.01	1.01	1.02
	140	1.43	1.44	1.43	1.42	1.17	1.15	1.17	1.18	1.00	1.02	1.02	1.02	0.99	1.00	1.00	1.01

$Accuracy\{se(\cdot)\}$ denotes the ratio of the estimated standard errors (of the average effects) to the sampling standard deviations. $Ratio\{se(\cdot)\}$ denotes the ratio of the estimated standard errors to the infeasible standard errors. Results obtained using Option 2, when $E(\sigma_i^2) = 1$.

[10] The accuracy of the estimated standard errors is computed as the ratio of $B^{-1}\sum_{b=1}^{B}\left\{\sqrt{(\hat{\Phi}_b)_{kk}}\right\}$ to the sampling standard deviation of $\hat{\psi}_k$, given by the square root of $(B-1)^{-1}\sum_{b=1}^{B}\left(\hat{\psi}_{k,(b)} - \bar{\hat{\psi}}_k\right)^2$, where $\bar{\hat{\psi}}_k = B^{-1}\sum_{b=1}^{B}\hat{\psi}_{k,(b)}$, for $k = 1, 2$.

Results reported in Table 4, show that relying exclusively on the asymptotic properties of $\hat{\triangle}_1$ may lead to invalid inference in finite samples. The estimated standard errors are upwards biased for the vast majority of cases. These biases can be substantial unless T and N or the degree of coefficient heterogeneity (σ_β) are large. However, if the coefficient dispersion is low, the estimated standard errors can be largely overestimated even when both T and N are equal to 140. These biases can in turn significantly affect inference.

4.4.2 Hypothesis tests

To test the hypothesis $\psi = \underline{\psi}$, for $\underline{\psi}$ a known $K \times 1$ vector, Swamy (1970) suggests the following criterion:

$$F\left(\hat{\psi}, \underline{\psi}, \hat{\Phi}\right) = \frac{N-K}{K\,(N-1)} \left(\hat{\psi} - \underline{\psi}\right)' \hat{\Phi}^{-1} \left(\hat{\psi} - \underline{\psi}\right). \tag{24}$$

The asymptotic distribution of the test is F, with K, $N-K$ degrees of freedom.

Empirical Moments of the F-statistic. We now study the finite-sample properties of the distribution of (24). In particular, we examine the empirical distributions of $F\left(\hat{\psi}_{k,GLS}, \psi_{0,k}, \Phi_{kk}\right)$, and $F\left(\hat{\psi}_{k,FGLS}, \psi_{0,k}, \hat{\Phi}_{kk}\right)$, computed under the null hypothesis that the estimator (of interest) of ψ_k is equal to the corresponding true value used to generate the data, $\psi_{0,k}$, for $k = 1, 2$.[11] We then compare the mean, standard deviation, skewness, and excess kurtosis of these two empirical distributions with the corresponding population moments of a F-distribution with $1, N-1$ degrees of freedom. Results are reported in Table 5.

In many cases, the means and standard deviations of the distributions of the F-statistics based on the infeasible GLS estimator are relatively close to the true means and standard deviations. This is not the case when considering the distributions of the F-statistics based on the feasible GLS estimator. The means and standard deviations of the latter can be substantially smaller than the values associated with a F-distribution with 1, $N-1$ degrees of freedom. Results worsen when testing hypothesis about the intercept rather than slope parameters. These results are in line with the fact that $\hat{\triangle}_1$ is often upwards biased. The skewness and excess-kurtosis of the distribution of the F-statistics based on both feasible and infeasible GLS estimators, can be far from the corresponding population moments unless N is large.

[11] Φ_{kk} is the kth diagonal element of Φ. Similarly, ψ_k denotes the kth element of ψ.

Table 5: Empirical moments of F-statistics

$\boldsymbol{\beta}$		Mean				Standard Deviation				Skewness				Excess-Kurtosis			
	$T\backslash N$	10	30	50	140	10	30	50	140	10	30	50	140	10	30	50	140
$F_{1,N-1}$		**1.29**	**1.07**	**1.04**	**1.01**	**2.30**	**1.61**	**1.52**	**1.45**	**6.71**	**3.37**	**3.12**	**2.92**	**214.50**	**19.23**	**15.61**	**13.11**
Infeasible	10	0.96	1.06	0.97	1.04	1.37	1.47	1.47	1.47	2.86	2.40	3.17	2.60	11.67	7.23	14.87	8.78
	20	1.01	1.00	0.94	0.99	1.47	1.56	1.35	1.41	2.92	3.65	2.86	2.95	12.66	21.85	11.26	12.12
	30	**1.02**	**0.98**	**1.03**	**0.97**	**1.38**	**1.42**	**1.40**	**1.41**	**2.44**	**2.65**	**2.35**	**2.92**	**7.69**	**9.43**	**7.06**	**12.41**
	50	1.08	0.96	0.96	1.06	1.58	1.33	1.33	1.53	3.06	2.65	2.52	2.73	13.51	9.55	8.06	9.99
	70	1.00	1.02	0.98	0.98	1.42	1.47	1.46	1.49	3.74	3.41	3.32	3.12	28.41	19.85	17.49	14.11
	140	0.96	0.96	1.07	1.02	1.31	1.33	1.43	1.44	2.64	2.61	2.58	2.51	9.77	8.82	9.14	7.56
Feasible	10	0.37	0.34	0.39	0.39	0.56	0.48	0.55	0.54	3.48	2.95	2.77	2.61	18.91	13.18	11.11	10.25
	20	0.55	0.46	0.44	0.43	0.86	0.72	0.63	0.64	3.86	3.42	2.90	3.24	23.48	17.91	12.25	15.69
	30	**0.58**	**0.48**	**0.51**	**0.48**	**0.85**	**0.70**	**0.69**	**0.69**	**3.25**	**3.59**	**2.49**	**2.97**	**15.13**	**26.34**	**8.97**	**12.46**
	50	0.71	0.56	0.56	0.59	1.14	0.81	0.79	0.85	3.59	3.04	2.74	2.83	18.00	13.08	10.30	11.52
	70	0.70	0.64	0.62	0.62	1.09	0.97	0.94	0.96	4.45	4.48	3.34	3.31	37.08	40.21	17.46	16.76
	140	0.79	0.76	0.82	0.75	1.27	1.11	1.14	1.06	4.36	3.06	2.74	2.56	34.97	13.97	10.30	8.02

$\boldsymbol{c}$		Mean				Standard Deviation				Skewness				Excess-Kurtosis			
	$T\backslash N$	10	30	50	140	10	30	50	140	10	30	50	140	10	30	50	140
$F_{1,N-1}$		**1.29**	**1.07**	**1.04**	**1.01**	**2.30**	**1.61**	**1.52**	**1.45**	**6.71**	**3.37**	**3.12**	**2.92**	**214.50**	**19.23**	**15.61**	**13.11**
Infeasible	10	1.09	1.01	0.96	1.01	1.43	1.37	1.40	1.43	2.34	2.50	2.74	2.67	7.00	8.57	9.76	9.27
	20	0.99	1.00	1.02	0.97	1.40	1.43	1.41	1.44	2.87	2.94	2.72	4.32	12.95	11.53	10.39	38.22
	30	**1.02**	**0.95**	**0.99**	**1.01**	**1.45**	**1.37**	**1.40**	**1.38**	**2.56**	**3.05**	**2.73**	**2.73**	**8.44**	**15.32**	**10.26**	**11.25**
	50	1.02	1.00	0.98	0.98	1.54	1.36	1.37	1.40	2.87	2.64	2.53	3.03	10.30	11.47	8.26	14.45
	70	0.99	1.05	0.95	1.00	1.49	1.50	1.38	1.43	3.17	2.52	3.11	3.07	15.04	7.93	14.63	15.15
	140	0.99	1.01	1.00	0.99	1.35	1.50	1.43	1.34	2.43	2.93	2.78	2.33	7.81	11.53	10.54	6.71
Feasible	10	0.38	0.22	0.31	0.29	0.59	0.30	0.43	0.43	4.76	2.47	2.55	2.77	45.68	8.11	8.78	10.07
	20	0.46	0.35	0.34	0.33	0.77	0.53	0.49	0.47	5.41	3.53	2.79	3.14	51.13	19.04	10.70	14.86
	30	**0.43**	**0.32**	**0.38**	**0.36**	**0.62**	**0.44**	**0.55**	**0.49**	**2.69**	**2.93**	**2.76**	**2.67**	**9.45**	**13.93**	**11.42**	**10.02**
	50	0.47	0.42	0.40	0.40	0.70	0.60	0.57	0.58	3.11	3.05	2.80	3.19	13.65	13.79	11.90	14.99
	70	0.53	0.49	0.44	0.44	0.81	0.71	0.65	0.62	3.72	2.95	3.05	2.90	22.65	12.36	13.24	13.28
	140	0.57	0.57	0.54	0.53	0.86	0.92	0.78	0.73	3.27	4.21	2.85	2.52	16.53	31.53	11.26	8.51

Empirical moments of F-statistics across different sample sizes $(T.N)$, when the data are generated from Option 2, with $E(\sigma_i^2) = 1$, and $\sigma_\beta = \sigma_c = 0.1$. In the upper panel, the test statistics are constructed under the null hypothesis H_0: $\beta = 0.5$ against the alternative H_1: $\beta \neq 0.5$. In the lower panel, the null hypothesis is H_0: $c = 0$ against H_1: $c \neq 0$. Row "$F_{1,N-1}$" reports the population moments of a F-distribution with 1, $N-1$ degrees of freedom. The empirical moments reported in "Infeasible" correspond to the F-statistics computed using the infeasible GLS estimator of ψ and the infeasible covariance matrix, Φ. "Feasible" is used to denote the empirical moments of the F-statistics, replacing the unknown components in ψ and Φ by their estimators.

Power Performances. In Table 6 we report the empirical sizes of the F-statistic, described in equation (24), of the null hypothesis H_0: $\psi_k = \psi_{0,k}$ against the alternative H_1: $\psi_k \neq \psi_{0,k}$. They are computed as the relative rejection frequencies based on the critical regions of nominal size 0.05 of a F-distribution with 1, $N-1$ degrees of freedom. This allows us to evaluate the direct consequences of the various results described above for hypothesis tests.

Table 6: Empirical sizes based on F-statistics

		$\sigma_\beta = 0.05$				$\sigma_\beta = 0.1$				$\sigma_\beta = 0.3$				$\sigma_\beta = 0.5$			
	$T\backslash N$	10	30	50	140	10	30	50	140	10	30	50	140	10	30	50	140
$size\left(\hat{\beta}_{GLS}\right)$	10	2.47	4.80	4.47	**4.00**	2.13	5.07	4.60	**5.33**	2.13	4.00	4.60	**5.67**	2.27	3.80	4.00	**4.13**
	20	2.67	4.53	4.40	**5.67**	2.93	4.40	3.73	**3.73**	2.80	4.40	4.00	**5.27**	2.47	3.27	3.47	**4.80**
	30	2.20	4.73	4.60	**4.60**	2.33	4.33	5.13	**4.67**	1.60	4.47	3.13	**4.87**	2.00	3.87	4.47	**5.33**
	50	2.27	3.87	3.80	**5.40**	3.07	3.53	4.00	**5.47**	2.60	5.00	3.40	**4.53**	2.20	4.07	4.00	**5.07**
	70	2.33	3.33	3.93	**4.27**	1.87	3.53	4.13	**5.07**	2.60	4.33	5.27	**4.07**	2.33	3.67	5.27	**3.80**
	140	2.53	3.33	4.13	**4.53**	2.13	3.47	4.80	**5.80**	1.93	3.60	4.40	**4.93**	3.33	4.27	4.87	**4.73**
$size\left(\hat{\beta}_{FGLS}\right)$	10	0.00	0.07	0.07	**0.20**	0.13	0.07	0.13	**0.07**	1.07	0.60	0.73	**0.60**	2.20	1.00	1.67	**2.00**
	20	0.07	0.20	0.00	**0.13**	0.40	0.47	0.27	**0.33**	1.80	2.47	2.13	**2.27**	3.47	2.93	2.93	**3.60**
	30	0.20	0.27	0.13	**0.13**	0.67	0.20	0.27	**0.53**	2.20	3.20	2.13	**3.00**	2.73	3.47	3.60	**4.40**
	50	0.33	0.07	0.20	**0.60**	1.20	0.93	0.93	**1.07**	3.07	4.53	3.07	**3.00**	3.87	4.47	3.93	**4.53**
	70	0.20	0.27	0.53	**0.33**	0.87	0.87	1.27	**1.40**	2.93	4.00	4.87	**3.40**	4.60	4.67	4.47	**3.73**
	140	0.40	0.47	0.47	**0.53**	1.33	2.13	2.40	**2.47**	4.33	4.00	4.60	**4.87**	5.33	5.60	5.33	**4.33**
$size\left(\hat{c}_{GLS}\right)$	10	2.33	4.00	4.33	**4.27**	2.53	3.93	3.80	**4.53**	1.93	3.87	3.47	**4.40**	2.33	3.47	3.67	**5.33**
	20	2.47	4.53	3.93	**4.27**	2.13	3.87	3.93	**4.07**	3.20	4.33	5.27	**4.27**	3.00	3.60	4.00	**5.53**
	30	2.60	5.00	3.80	**5.13**	2.73	3.60	4.13	**5.13**	1.67	4.27	4.07	**3.93**	2.27	3.87	4.33	**6.60**
	50	2.60	4.00	4.53	**4.13**	3.13	3.80	4.40	**4.80**	2.13	4.80	3.93	**5.33**	2.33	3.07	3.53	**5.13**
	70	2.00	4.40	5.27	**4.53**	2.53	4.80	3.80	**4.60**	3.00	4.20	3.00	**3.73**	2.27	3.20	4.87	**4.07**
	140	2.07	4.80	3.47	**4.00**	2.07	4.27	4.13	**4.67**	1.93	4.13	4.73	**4.47**	2.87	5.33	5.00	**4.73**
$size\left(\hat{c}_{FGLS}\right)$	10	0.07	0.07	0.00	**0.00**	0.07	0.00	0.00	**0.00**	0.20	0.27	0.47	**0.20**	1.20	0.40	0.53	**0.53**
	20	0.13	0.07	0.13	**0.07**	0.40	0.20	0.07	**0.07**	0.73	0.53	1.27	**0.67**	1.73	1.47	1.80	**2.07**
	30	0.07	0.13	0.13	**0.07**	0.07	0.07	0.07	**0.13**	0.73	1.40	1.20	**0.60**	1.80	2.33	2.60	**3.53**
	50	0.07	0.07	0.13	**0.00**	0.20	0.27	0.20	**0.27**	1.67	1.73	1.80	**1.80**	2.13	2.27	2.60	**3.00**
	70	0.00	0.07	0.13	**0.13**	0.33	0.47	0.33	**0.27**	2.60	1.87	1.93	**1.93**	3.53	2.40	3.80	**3.00**
	140	0.27	0.13	0.27	**0.20**	0.33	1.20	0.73	**0.47**	3.00	3.47	3.00	**3.60**	5.27	5.47	5.00	**3.87**

Rejection frequencies (%) at 5% nominal level obtained computing the F-statistic described in (24), under the null hypothesis H_0: $\psi_k = \psi_{0,k}$ against the alternative H_1: $\psi_k \neq \psi_{0,k}$. $\hat{\beta}_{GLS}$ and $\hat{c}_{GLS}$ denote the infeasible GLS estimator of β and c respectively. Similarly, the subscript "FGLS" stands for feasible GLS. The data are generated from Option 2, with $E(\sigma_i^2) = 1$.

The tests based on the feasible GLS estimation severely suffer from size distortions. Unless the degree of coefficient heterogeneity is quite high (e.g. $\sigma_\beta = 0.5$), the sizes are always substantially lower than the nominal levels. They are often close to zero due to the fact that the estimated standard errors are largely biased upward. Once again, the distortions are even more severe when testing about the intercept parameters.

To support these findings, we plot the power functions for the slope and intercept parameters in Figure 1 and 2 respectively. To save space, we only report results for the case with $E(\sigma_i^2) = 1$ and $\sigma_\beta = 0.1$.[12]

[12] Additional results are available in an online Appendix.

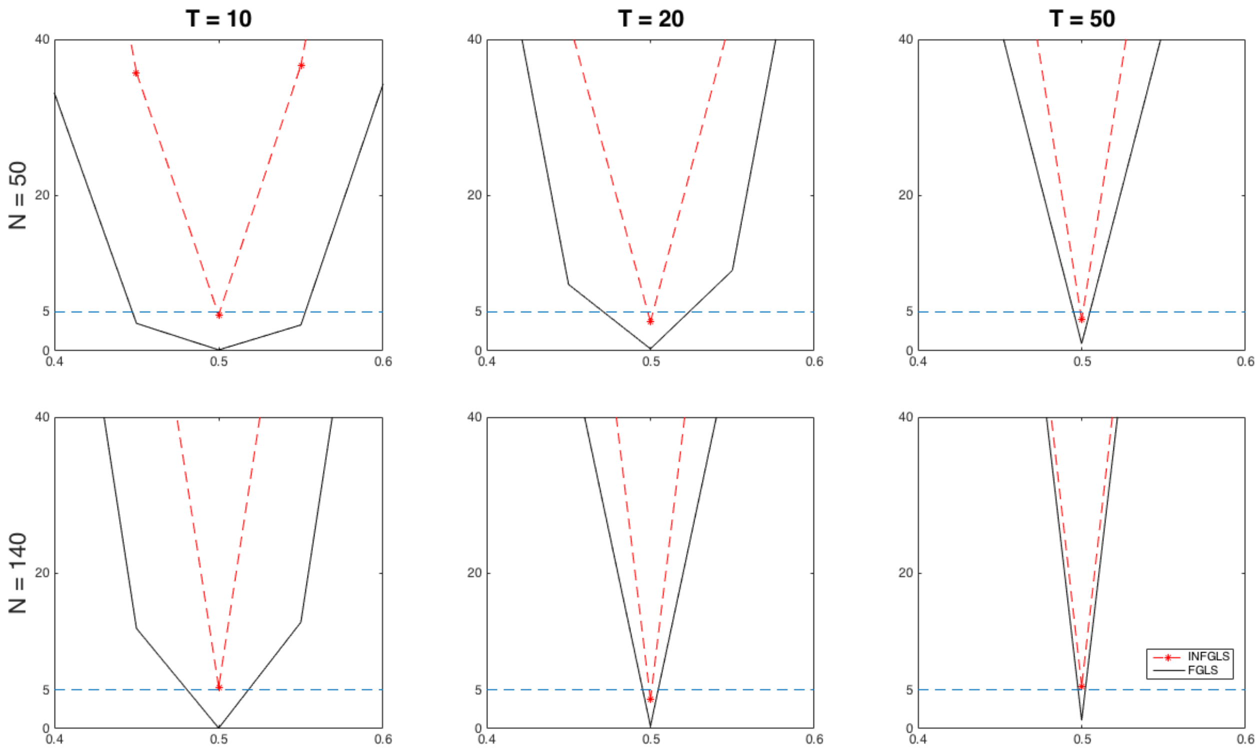

Figure 1: Rejection frequency (%) at the 5% nominal level, for the slope parameter (β), in the y-axis. They are computed using the F-statistic described in (24), under the null hypothesis H_0: $\beta = \underline{\beta}$ against the alternative $\beta \neq \underline{\beta}$. Different values of $\underline{\beta}$ are reported in the x-axis. The true value of β is 0.5. The black lines and the red dotted lines denote the power performances of feasible and infeasible GLS estimators, respectively. Results obtained using Option 2, with $E(\sigma_i^2) = 1$ and $\sigma_\beta = 0.1$.

5 Conclusions

As in the error component model, the estimator of the coefficients' covariance matrix in a random coefficient model is often negative definite. The aim of this study is to investigate the causes and effects of the problem. By running some Monte Carlo experiments, we show that the degree of coefficient heterogeneity relative to the (conditional) variability of the dependent variables plays a crucial role. The larger the coefficient dispersion and the precision of the regression disturbances (the inverse of the average standard deviations of the time-varying errors), the lower the probability to observe a negative definite estimator of the random coefficient covariance matrix. An increase in the former has a larger effect than an increase in the latter. Similarly, this probability decreases as the time dimension and the number of units get large, partly due to the fact that the performances (in terms of bias and RMSE) of

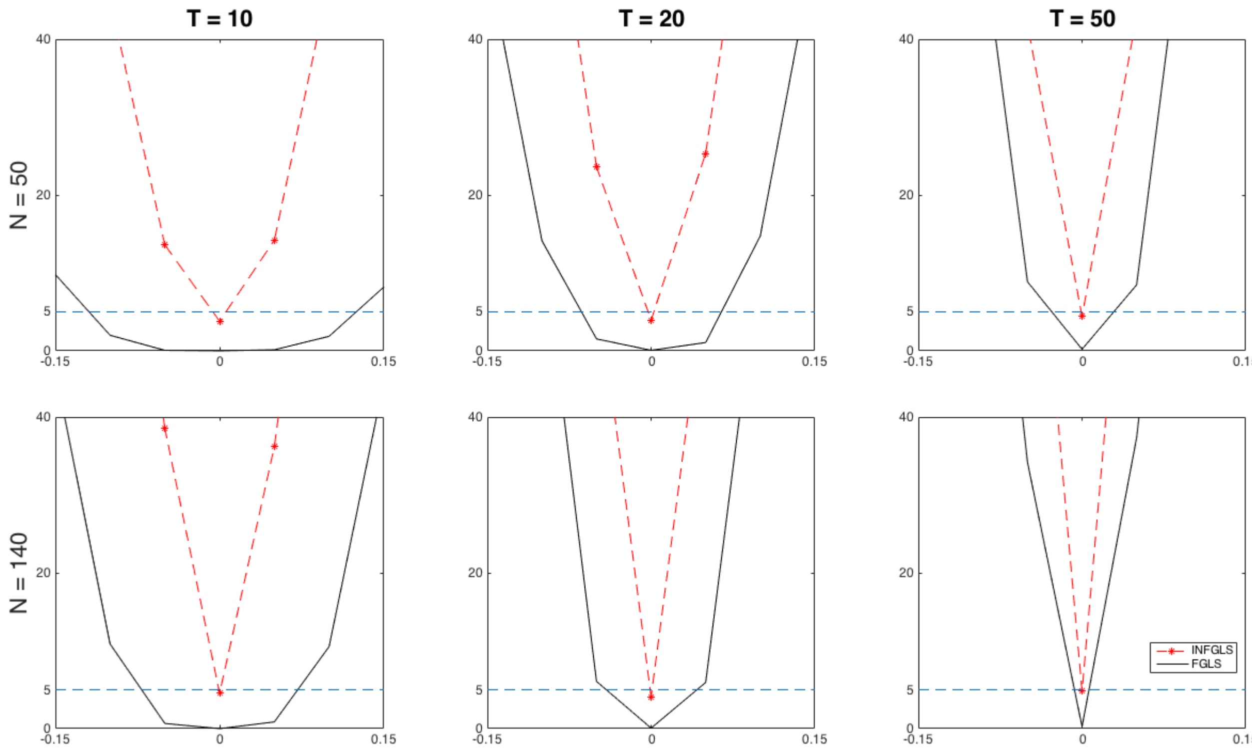

Figure 2: Rejection frequency (%) at the 5% nominal level, for the intercept parameter (c), in the y-axis. They are computed using the F-statistic described in (24), under the null hypothesis H_0: $c = \underline{c}$ against the alternative $c \neq \underline{c}$. Different values of $\underline{c}$ are reported in the x-axis. The true value of c is 0. The black lines and the red dotted lines denote the power performances of feasible and infeasible GLS estimators, respectively. Results obtained using Option 2, with $E(\sigma_i^2) = 1$ and $\sigma_c = 0.1$.

individual OLS estimates and the Mean Group improves in large samples. It is known that when the time dimension goes to infinity, the negative definiteness problem vanishes.

We then demonstrate that relying on the asymptotic properties of the biased but consistent estimator of the random coefficient covariance matrix may lead to poor finite-sample inference. Unless the time and cross-section dimensions, and/or the degree of coefficient dispersion are high, the estimated standard errors are largely upwards biased. The resulting hypothesis tests may suffer from considerable size distortions. The empirical sizes of the tests are substantially lower than the nominal levels. Results may worsen when the precision of the regression disturbances decreases. An estimation procedure which yields an unbiased and more efficient estimator of the random coefficient covariance and which performs relatively well in terms of accuracy of inference is being proposed in a separate paper.

A Appendix

A.1 Estimation of Ω_i when the disturbances exhibit serial correlation

Swamy (1971) considers the estimation problem of Ω_i when the the disturbances follow an AR(1) process:

$$u_{it} = \phi_i u_{i,t-1} + \epsilon_{it}, \qquad 0 <\mid \phi_i \mid< 1, \tag{25}$$

and $E\left(\epsilon_{it}\right) = 0$, $E\left(\epsilon_{it}\epsilon_{js}\right) = \sigma_i^2$ if $t = s$ and $i = j$, and 0 otherwise. For $i = j$, $E\left(u_i u_j'\right) = \sigma_i^2 \Omega_i$, where

$$\Omega_i = \frac{1}{1-\phi_i^2}\begin{bmatrix} 1 & \phi_i & \phi_i^2 & \cdots & \phi_i^{T-1} \\ \phi_i & 1 & \phi_i & \cdots & \phi_i^{T-2} \\ \phi_i^2 & \phi & 1 & & \phi_i^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_i^{T-1} & \phi_i^{T-2} & \phi_i^{T-3} & \cdots & 1 \end{bmatrix},$$

and $E\left(u_i u_j'\right) = 0$, if $i \neq j$. A consistent estimator of ϕ_i is given by

$$\hat{\phi}_i = \frac{\sum_{t=2}^{T} \hat{u}_{it}\hat{u}_{i,t-1}}{\sum_{t=2}^{T} \hat{u}_{i,t-1}^2}, \tag{26}$$

where $\hat{u}_{it}$ is the t-th element of $\hat{u}_i$, the vector of OLS residuals. An estimator of Ω_i can be obtained by replacing ϕ_i by $\hat{\phi}_i$ in Ω_i. Note also that the inverse of $\hat{\Omega}_i$ can be computed using the fact that $\hat{\Omega}_i^{-1} = \hat{R}_i'\hat{R}_i$, where

$$\hat{R}_i = \begin{bmatrix} \sqrt{1-\hat{\phi}_i^2} & 0 & 0 & \cdots & 0 & 0 \\ -\hat{\phi}_i & 1 & 0 & & & 0 \\ 0 & -\hat{\phi}_i & 1 & & & \vdots \\ \vdots & & \ddots & \ddots & & \vdots \\ & & & & & \\ 0 & 0 & \cdots & & -\hat{\phi}_i & 1 \end{bmatrix}.$$

Under assumption (25), by regressing y_i upon X_i, and applying Aitken's GLS to each time series, we have

$$\begin{aligned} \hat{\beta}_{i,gls} &= (X_i'\Omega_i^{-1}X_i)^{-1}X_i'\Omega_i^{-1}y_i \\ &= \beta + \delta_i + (X_i'\Omega_i^{-1}X_i)^{-1}X_i'\Omega_i^{-1}u_i. \end{aligned} \tag{27}$$

The feasible GLS estimator of β_i is given by

$$\tilde{\beta}_i = (X_i'\hat{\Omega}_i^{-1}X_i)^{-1}X_i'\hat{\Omega}_i^{-1}y_i. \tag{28}$$

The average effect, β, can be estimated by

$$\hat{\beta}_{FGLS} = \sum_{i=1}^{N} \hat{W}_i\tilde{\beta}_i, \tag{29}$$

where

$$\hat{W}_i = \left\{\sum_{i=1}^{N}\left[\hat{\triangle} + \tilde{\sigma}_i^2(X_i'\hat{\Omega}_i^{-1}X_i)^{-1}\right]^{-1}\right\}^{-1}\left[\hat{\triangle} + \tilde{\sigma}_i^2(X_i'\hat{\Omega}_i^{-1}X_i)^{-1}\right]^{-1},$$

$$\tilde{\sigma}_i^2 = \frac{\tilde{u}_i'\tilde{u}_i}{T-K},$$

and $\tilde{u}_i = \hat{R}_iy_i - \hat{R}_iX_i\tilde{\beta}_i$.

Similarly to (12), using (27), we can compute

$$var\left(\hat{\beta}_{i,gls}\right) = E\left(\hat{\beta}_{i,gls} - \beta\right)\left(\hat{\beta}_{i,gls} - \beta\right)' = \triangle + \sigma_i^2(X_i'\Omega_i^{-1}X_i)^{-1}.$$

The estimator of $\triangle$ becomes

$$\hat{\triangle} = \hat{\triangle}_4 - \hat{\triangle}_5, \tag{30}$$

where

$$\begin{aligned} \hat{\triangle}_4 &= \tfrac{1}{N-1}\textstyle\sum_{i=1}^{N}\left(\tilde{\beta}_i - N^{-1}\sum_{i=1}^{N}\tilde{\beta}_i\right)\left(\tilde{\beta}_i - N^{-1}\sum_{i=1}^{N}\tilde{\beta}_i\right)', \\ \hat{\triangle}_5 &= \tfrac{1}{N}\textstyle\sum_{i=1}^{N}\tilde{\sigma}_i^2\left(X_i'\hat{\Omega}_iX_i\right)^{-1}. \end{aligned} \tag{31}$$

References

[1] B. Baltagi. *Econometric analysis of panel data.* John Wiley & Sons, 2008.

[2] B. H. Baltagi, G. Bresson, J. M. Griffin, and A. Pirotte. Homogeneous, heterogeneous or shrinkage estimators? Some empirical evidence from french regional gasoline consumption. *Empirical Economics*, 28(4):795–811, 2003.

[3] B. H. Baltagi and J. M. Griffin. Pooled estimators vs. their heterogeneous counterparts in the context of dynamic demand for gasoline. *Journal of Econometrics*, 77(2):303–327, 1997.

[4] B. H. Baltagi, J. M. Griffin, and W. Xiong. To pool or not to pool: Homogeneous versus heterogeneous estimators applied to cigarette demand. *Review of Economics and Statistics*, 82(1):117–126, 2000.

[5] D. Boyd and R. P. Smith. Some econometric issues in measuring the monetary transmission mechanism, with an application to developing countries. In L. Mahadeva and P. Sinclair, editors, *Monetary transmission in diverse economies*, pages 68–99. Cambridge University Press, 2002.

[6] H. Brücker and B. Siliverstovs. On the estimation and forecasting of international migration: how relevant is heterogeneity across countries? *Empirical Economics*, 31(3):735–754, 2006.

[7] A.-M. Fuertes and R. Smith. *Panel Time Series*. CEMMAP, 2016.

[8] A. S. Goldberger. *Econometric theory.* New York: John Wiley & Sons, 1964.

[9] C. Hsiao. *Analysis of panel data*. Cambridge University Press, 2003.

[10] J. G. MacKinnon and H. White. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29(3):305–325, 1985.

[11] M. H. Pesaran and R. Smith. Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, 68(1):79–113, 1995.

[12] P. Swamy. *Statistical inference in random coefficient regression models*, volume 55. Springe-Verlag, 1971.

[13] P. A. Swamy. Efficient inference in a random coefficient regression model. *Econometrica: Journal of the Econometric Society*, pages 311–323, 1970.

[14] L. Trapani and G. Urga. Optimal forecasting with heterogeneous panels: A monte carlo study. *International Journal of Forecasting*, 25(3):567–586, 2009.

[15] H. White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, pages 817–838, 1980.